Manuscript No. TGRS - 2022 - 03317

Forecasting Earthquake Magnitude and Epicenter by Incorporating Spatio-temporal Priors into Deep Neural Networks

Jie Liu, Tong Zhang, Chulin Gao and Peixiao Wang

***Abstract*—Forecasting earthquake magnitude and epicenter is of great significance for disaster management and hazard mitigation. Existing machine learning-based earthquake forecasting has faced two shortcomings: limited historical earthquake samples for training and lack of explicit consideration of seismic prior knowledge. We propose a novel Spatio-Temporal Priors-Informed Deep Networks (STPiDN) that incorporates seismic spatio-temporal prior knowledge into deep neural networks using limited historical earthquake samples. In our method, a Physics-informed Recurrent Graph Network (PRGN) is developed to extract representations of observed earthquake precursor data in a physics-informed manner. In order to make prior-guided earthquake predictions, we develop seismic prior knowledge activation layers that combine earthquake event representations with prior knowledge (e.g. fault distribution) through activation gates. An adaptive multi-task loss function is proposed to achieve joint magnitude and epicenter forecasting with the consideration of alleviating the magnitude and epicenter imbalance problem. Our empirical evaluation results show that the proposed forecasting method outperforms several competing methods on a real-world earthquake dataset, proving that physics-informed prediction methods have the potential to capture complex earthquake patterns using limited training samples.**

***Index Terms*—Earthquake forecasting, epicenter, physics-informed recurrent graph network, seismic prior knowledge**

I. Introduction

EARTHQUAKES are formidable natural disasters that can cause massive human casualties and economic losses. Timely earthquake prediction is of considerable importance for disaster response and management [1]–[3] and has attracted extensive research attention [4], [5]. The prediction of future earthquake occurrence, magnitude and epicenter is of great significance for effective earthquake hazard mitigation. Most existing studies have focused on occurrence prediction [6], [7] or magnitude prediction [8]–[10]. There is little research on the joint prediction of earthquake occurrence, magnitude and epicenter.

Statistical methods such as the aftershock-sequence model [11] and probabilistic analysis [12] have been widely used for earthquake prediction. However, statistical methods usually require prior knowledge of data distribution, which is difficult to determine [12]. Advanced Machine Learning (ML) techniques have been used in earthquake engineering [13], [14], such as ground motion prediction using artificial neural network [15] and the control of seismic structural control system via reinforcement learning [16]. Acoustic & Electromagnetic Testing All in one system (AETA) team has been working on earthquake prediction for more than 10 years. Starting from designing sensors and building observation stations, the AETA team has gradually developed the AETA dataset for earthquake prediction studies [17]. Based on the AETA dataset, machine learning techniques have been used for earthquake prediction and achieved promising results. Principle component analysis was used to extract the main components of the low electromagnetic feature from multiple stations and light gradient boosting machine (LGB) was used to predict the occurrence of earthquakes [18]. Convolutional Neural Network (CNN) was used for earthquake magnitude prediction based on AETA 3D feature maps and outperformed classical computer vision methods(e.g. Resnet and VGG) [19].

In the literature, ML-driven earthquake prediction methods, such as random forests [8], [20], logistic regression [21], artificial neural network [22], [23] and support vector machine [24], [25], have been applied for earthquake prediction due to their capabilities to capture hidden patterns of earthquake occurrence [26]. More recently, deep learning methods, such as deep neural networks [4], [27], CNN [28], [29] and long-short term memory [30], [31] have gained momentum. Some recent earthquake prediction methods used seismic data from one single station, such as the regression of earthquake location using Bayesian neural networks [32] and the prediction of epicentral distance, depth, & magnitude using complex neural networks [33]. These two studies predict earthquakes within 350km of one single station. In these methods, seismic prior knowledge is mainly used implicitly in the form of model inputs (e.g. [6], [22]), which may not be fully utilized in a physics-informed manner. In addition, the fault distribution priors and the physical laws of earthquakes are not adequately considered in these ML methods [6], [8].

The challenges and shortcomings of current ML earthquake prediction methods are summarized as follows: 1) Based on limited historical training samples, it is difficult to train robust predictors to generalize occurrence patterns and prior

Manuscript received September 9, 2022. This work was supported in part by the joint Research Project for Meteorological Capacity Improvement under grant 22NLTSY015 and Hubei Provincial Natural Science Foundation of China under Grant 2022CFD012. (Corresponding author: Tong Zhang.).

Jie Liu, Tong Zhang, Chulin Gao and Peixiao Wang are with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China (e-mail: liu_whu@whu.edu.cn; zhangt@whu.edu.cn; clgao@whu.edu.cn; peixiaowang@whu.edu.cn).

knowledge. 2) They usually do not explicitly account for seismic prior knowledge, such as fault distribution, historical earthquake representations, and physical laws, which may prevent these methods from producing accurate and robust results. For example, the magnitude and epicenter of an earthquake are closely related to the nearby fault activities because earthquakes are usually caused by the stress accumulation and rupture of faults [34].

To alleviate the above problems, we propose a multi-task deep learning method, called Spatio-Temporal Priors-Informed Deep Networks (STPiDN), which integrates spatio-temporal seismic priors to predict the magnitude and epicenter of earthquakes. Priors mainly include wave equation, spatial distribution of faults, conditional probability of earthquake recurrence and historical earthquake representations, which are incorporated in the core components of STPiDN: a physics-informed recurrent graph network (PRGN) and seismic prior knowledge activation layers. Using the physical law of wave equation as priors, the PRGN embeds observed earthquake precursor data (e.g. electromagnetic & geoacoustic data provided by AETA) into the event representations for predicting future earthquakes. The seismic prior knowledge activation layers are used to incorporate prior knowledge (e.g. spatial distribution of faults, conditional probability of earthquake recurrence and historical earthquake representations) into the learned earthquake event representations. The entire earthquake prediction method is implemented in an end-to-end learning framework using an adaptive multi-task loss function to jointly predict the magnitude and epicenter of future earthquakes.

The contributions of this study are summarized as follows:

1) We present a Spatio-Temporal Priors-Informed Deep Networks (STPiDN) for predicting earthquake magnitude and epicenter using an end-to-end unified learning scheme that integrates seismic prior knowledge to model complex physical earthquake processes.

2) A Physics-informed Recurrent Graph Network (PRGN) for representing historical earthquake events is developed. Inspired by the physical law of wave equation, we design a novel update function to extract representations of observed earthquake precursor data in a physics-informed manner.

3) We develop seismic prior knowledge activation layers that combine earthquake event representations with prior knowledge through activation gates, which facilitates the extraction of useful information for earthquake prediction.

4) Evaluation results on a real-world earthquake dataset demonstrate that the proposed STPiDN can accurately forecast the occurrence of earthquakes with an F1 score of 0.778 and a minimum mean epicenter offset of 271km, outperforming several existing baselines.

II. Related Work

A. Earthquake Prediction

Existing earthquake prediction methods can be classified into statistical and ML methods. Typical statistical methods include epidemic-type aftershock-sequence [11] and probabilistic analysis models [12]. Assuming specific data distributions, these methods are difficult to determine the optimal data distribution because the occurrence patterns of earthquakes vary in space and time [12]. Recently, ML methods have been extensively applied for earthquake prediction by virtue of their potential capabilities to model earthquake patterns [26]. Typical ML-driven earthquake prediction methods include LGB [18], random forests [8], [20], logistic regression[21], artificial neural network [22], [23], support vector machine[24], [25]. and more recently, deep learning methods, such as deep neural networks [4], [27], CNN [28], [29] and long-short term memory [30], [31]. Seismic prior knowledge has been mainly used in ML-driven earthquake prediction in the form of inputs. For example, the b-value, which is the slope of the Gutenberg-Richter's law curve [35], has been used as an input to ML-driven earthquake prediction methods [6], [22]. However, the prior knowledge of fault distribution, earthquake representations, and the physical law of earthquakes has rarely been explicitly incorporated into ML methods. In this paper, we attempt to integrate more comprehensive spatio-temporal prior knowledge in deep neural networks, aiming to improve the accuracy of earthquake prediction.

In terms of the training data used in ML methods, researchers mainly use earthquake catalogs and indicators derived from earthquake catalogs [6], [22]. Some earthquake predictors use potential earthquake precursors, such as ionospheric precursors [36]. A few studies adopt a mixed data strategy by using both earthquake precursors and earthquake catalogs [37]. Following the mixed data strategy, the proposed method uses observed electromagnetic & geoacoustic data and historical earthquake data.

Some recent earthquake prediction methods used seismic data from one single station and predict earthquakes within 350km of one single station [32], [33]. In our study, we predict future earthquakes in a large region (900km×1300km) based on the data from 95 stations. The full consideration of spatio-temporal correlations of seismic data from multiple stations may help capture the potential precursor information of earthquakes, which may be beneficial to earthquake prediction tasks.

B. Deep Learning-driven Spatio-temporal Prediction

In the past few years, an increasing number of studies have used deep learning methods for spatio-temporal prediction, which have the advantage of capturing the hidden patterns inherent in spatio-temporal processes through an end-to-end learning framework. The current state-of-the-art deep learning–driven prediction methods use hybrid architectures that use convolutional or graph neural networks to capture spatial correlations and use recurrent neural networks to learn temporal dynamics (e.g., [38]–[41]). Deep learning methods may face some limitations in spatio-temporal prediction, such as the need for sufficient data to guarantee performance and generalization, the inability to satisfy physical constraints, and the difficulty in interpreting results [42], [43]. These issues have motivated many studies to develop physic-informed deep learning methods that incorporate prior knowledge as physical constraints to improve the robustness and physical consistency

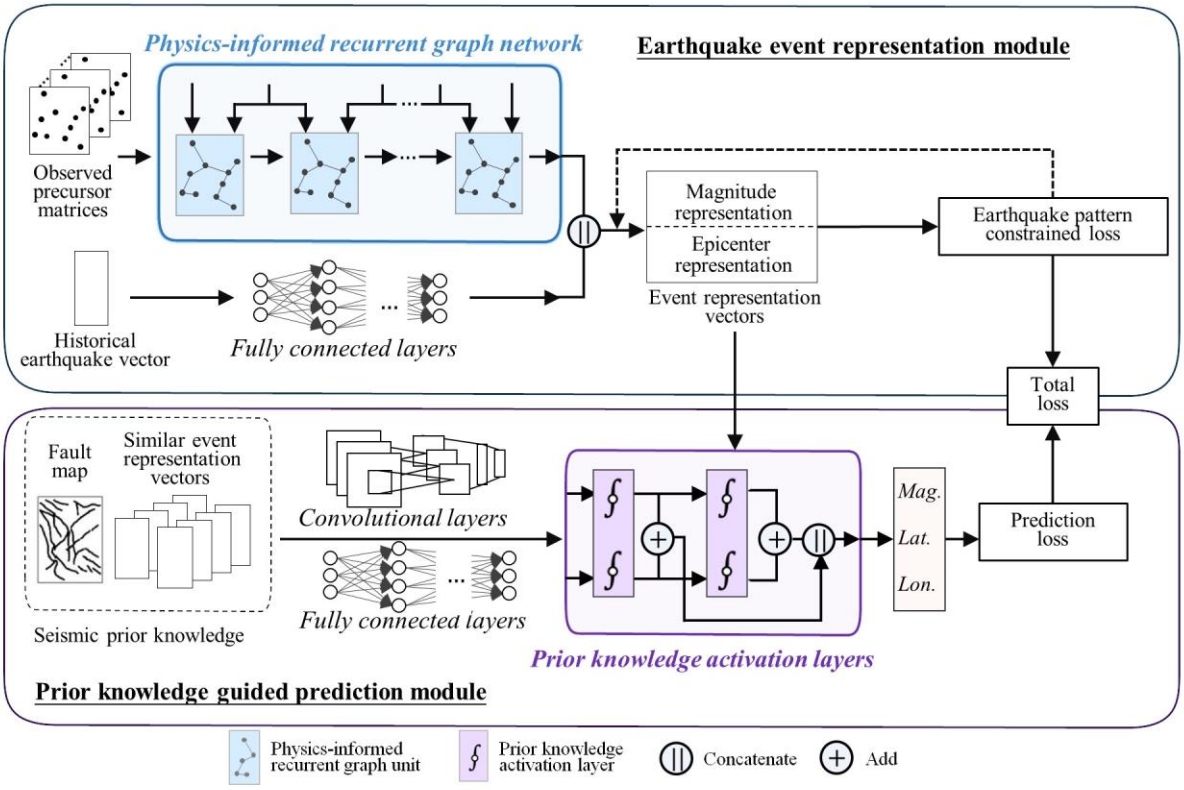


Fig. 1. The earthquake prediction method. 1) **The earthquake event representation module** produces event-oriented embeddings for each earthquake event using observed electromagnetic & geoacoustic data and historical earthquakes via the physics-informed recurrent graph network (PRGN) and a fully-connected neural network [4]. 2) **The prior knowledge-guided prediction module** predicts the magnitude and epicenter of earthquakes in the following week using multi-level activation features fused from earthquake representations and prior knowledge by seismic prior knowledge activation layers. Two modules are simultaneously optimized by an adaptive multi-task loss.

of conventional deep learning models [44], [45].

Prior knowledge can be explicitly incorporated into deep networks from the perspectives of training data, model architecture, loss function, and outputs [43]. Prior knowledge can be integrated as additional information sources of training data. For example, the b-value in the Gutenberg-Richter's law can be used as an input for earthquake prediction [6], [22]. Some researchers have used specific types of prior knowledge to design the architecture of deep learning network models [46], [47]. For example, Ling et al. [48] integrated invariant tensor basis from physical laws to embed Galilean invariance for the prediction of fluid anisotropy tensors. Chattopadhyay et al. [49] developed an equivariance-preserving deep spatial transformer for data-driven weather prediction. Prior knowledge can also be integrated into loss functions, such as the loss of energy conservation for lake temperature prediction [50] and the loss that describes the residuals of the governing partial differential equations for super-resolution of turbulent flows [51]. This study integrates comprehensive seismic prior knowledge (e.g. fault distribution, earthquake representations, and physical laws) into deep neural networks by developing novel network structures and loss functions.

III. Methodology

A. Problem Formulation and The Overall Framework

In this paper, our task is to predict the magnitude and epicenter of the largest earthquake (above magnitude 3.5) in the study region in the t+1 week using historical earthquakes and observed precursor data (electromagnetic & geoacoustic data) in t week. The earthquake prediction task can be formulated as,

$$\widehat{\boldsymbol{q}}_{\boldsymbol{t+1}} = \mathcal{F}\left\{\phi(\boldsymbol{q}_{\boldsymbol{t}}^{1}, \dots, \boldsymbol{q}_{\boldsymbol{t}}^{n_q}),\ [\boldsymbol{O}_{\boldsymbol{1}}, \dots, \boldsymbol{O}_{\boldsymbol{l}}, \dots, \boldsymbol{O}_{\boldsymbol{7}}]_{\boldsymbol{t}}, \boldsymbol{pk}_{\boldsymbol{t}}\right\} \quad (1)$$

where $\widehat{\boldsymbol{q}}_{\boldsymbol{t+1}} \in \mathbb{R}^3$ represents the predicted vector of the *largest* earthquake's magnitude, epicenter longitude and latitude in the study region in t+1 week. $\boldsymbol{q}_{\boldsymbol{t}}^{n_q}$ represents the vector of the last n_qth earthquake's magnitude, epicenter longitude and latitude before the t+1 week. Note the occurrence time of the earthquake in $\boldsymbol{q}_{\boldsymbol{t}}^{n_q}$ may be before the t week because of low occurrence rate. $\phi(\cdot)$ is a function that extracts h-dimensional explicit and implicit features from the last n_q earthquakes. $\boldsymbol{O}_{\boldsymbol{l}} \in \mathbb{R}^{N \times k}$ is a matrix that records the k-dimensional observed precursor data of N observation stations on the lth day of the tth week. $\boldsymbol{pk}_{\boldsymbol{t}}$ consists of prior knowledge that can be available before the t+1 week**:** faults, conditional probability of earthquake recurrence and historical earthquake representations**.** The first two priors are processed as gridded data with a spatial resolution of 0.25°.

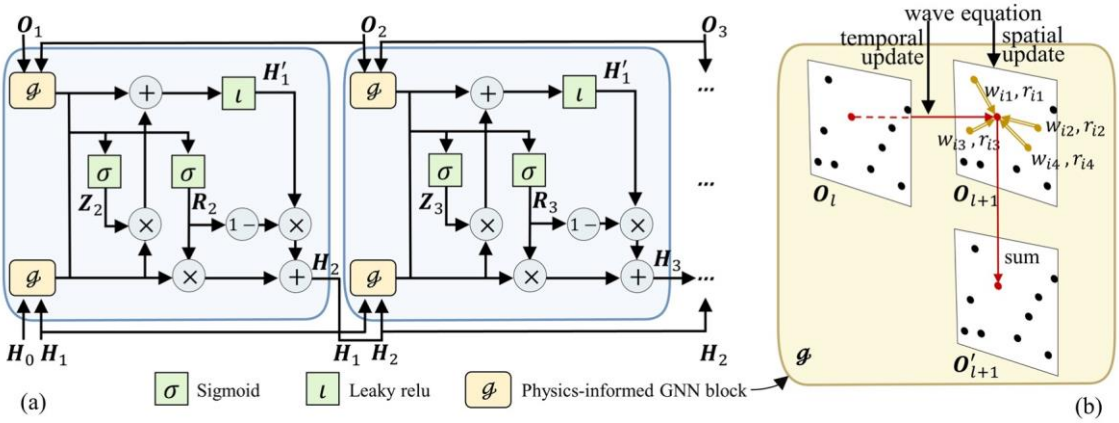


Fig. 2. Illustration of the physics-informed recurrent graph network. PRGN produces seismic embeddings from observed precursor data. (a) Each PRGU has two physics-informed GNN blocks to process observed precursor data $\boldsymbol{O}_l$and hidden memories $\boldsymbol{H}_l$, respectively. The updated memories are used as input for the next PRGU. (b) The output of Physics-informed GNN blocks (i.e. the updated node attributes $\boldsymbol{O}'_{l+1}$) is the weighted sum of spatial neighbours' attributes in $\boldsymbol{O}_{l+1}$and previous attributes in $\boldsymbol{O}_l$.

The last one is in vector form with a dimension of $n_e \times 64$. n_e is the number of used historical earthquake representations. Earthquake representation has a dimension of 64. $\mathcal{F}(\cdot)$ is a data-driven learnable prediction function. After obtaining the magnitude prediction, we can determine whether an earthquake with magnitude greater than 3.5 will occur in the study region in the following week.

The workflow of the proposed method is illustrated in Fig. 1. The method consists of two modules:

1) The earthquake event representation module produces event-oriented embeddings for each earthquake event using observed precursor data and historical earthquakes. The module uses an auxiliary task to divide the earthquake representation into magnitude representation and epicenter representation to facilitate the search for similar earthquake representations. The outputs of the module are vectors of magnitude and epicenter representations.

2) The prior knowledge-guided prediction module generates multi-level activation features by fusing earthquake representations and prior knowledge (e.g. fault distribution and historical earthquake representations). The module makes earthquake predictions using a two-layer fully connected neural network based on the generated multi-level activation features.

Two modules are simultaneously optimized by an adaptive multi-task loss function (i.e. an earthquake pattern constrained loss for the earthquake event representation module and a prediction loss for the prior knowledge guided prediction module).

The details of these two modules and the multi-task loss function are presented in sections III.B–III.D.

B. Earthquake Event Representation Module

In this module, the event representations are learned from observed electromagnetic & geoacoustic data from a sparse station network and historical earthquake data. Using sparse electromagnetic and geoacoustic data, we develop a Physics-informed Recurrent Graph Network (PRGN) to learn the representations of earthquake events. Guided by the wave equation that describes the relationship between electromagnetic waves and geoacoustic waves in time and space, the PRGN adaptively embeds the observed data while maintaining physical consistency. We use a seven-layer fully connected neural network [4] to compute the representations of earthquake events based on historical earthquake records (e.g. magnitude, epicenter and occurrence timestamp) and seismic indicators (e.g. b-value, seismic energy) calculated from earthquake catalogs. The representations from the two types of data (i.e., observed electromagnetic & geoacoustic data and earthquake catalog) are fused into an earthquake event representation vector $\boldsymbol{e}$ through two fully connected layers. It may be difficult to find earthquake event representations that are similar in both magnitude and epicenter. Therefore, each representation vector $\boldsymbol{e}$ is divided into a magnitude representation and an epicenter representation by an auxiliary constraint task to facilitate the search for more earthquakes with similar patterns in magnitude or epicenter.

To facilitate the embedding of observed electromagnetic & geoacoustic data, we define a graph $\boldsymbol{\mathcal{G}} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$ with N observation stations as its set of nodes $\boldsymbol{\mathcal{V}}$, and pairwise inter-correlations calculated by Gaussian distance decay-based functions [52] as its set of edges $\boldsymbol{\mathcal{E}}$. $\boldsymbol{v}_i^{(l)}, i = 1,2,\dots,N$ represents the ith node's attributes at time step l. Inspired by the gated graph neural network [53], PRGN adopts a hybrid architecture that uses Graph Neural Networks (GNN) to capture spatial-temporal correlations among observation stations and uses Gated Recurrent Units (GRU) to learn temporal dynamics of consecutive time steps. The update function of GNN incorporates the attribute information of stations from three consecutive time steps based on the wave propagation law. If the input of the GNN block is observed precursor data, the updated attribute information is electromagnetic & geoacoustic data; If the input of GNN block is hidden memory of GRU, the updated attribute information is the hidden state of GRU. The pipeline of PRGN is shown in Fig. 2, and the computation of observation data embeddings can be written as follows,

$$\begin{cases} \boldsymbol{Z}_{l+1} = \sigma(\mathcal{G}_{oz}[\boldsymbol{O}_l, \boldsymbol{O}_{l+1}] + \mathcal{G}_{hz}[\boldsymbol{H}_{l-1}, \boldsymbol{H}_l]) \\ \boldsymbol{R}_{l+1} = \sigma(\mathcal{G}_{or}[\boldsymbol{O}_l, \boldsymbol{O}_{l+1}] + \mathcal{G}_{hr}[\boldsymbol{H}_{l-1}, \boldsymbol{H}_l]) \\ \boldsymbol{H}'_{l+1} = \iota(\mathcal{G}_{oh}[\boldsymbol{O}_l, \boldsymbol{O}_{l+1}] + \boldsymbol{R}_{l+1} \circ \mathcal{G}_{hh}[\boldsymbol{H}_{l-1}, \boldsymbol{H}_l]) \\ \boldsymbol{H}_{l+1} = (1 - \boldsymbol{Z}_{l+1}) \circ \boldsymbol{H}'_{l+1} + \boldsymbol{Z}_{l+1} \circ \boldsymbol{H}_l \end{cases} \quad (2)$$

where $\boldsymbol{O}_l$ and $\boldsymbol{H}_l$ represent the observed data feature matrix and

hidden memories at time step *l*, respectively. $\boldsymbol{Z}_{l+1}$ is an update gate that determines which previous memories are kept in the current time step, and $\boldsymbol{R}_{l+1}$ is a reset gate that determines how new inputs are combined into the current memory. $\mathcal{g}(\cdot)$ represents a GNN block derived from the physical law of wave equation. The subscript letters o, z, h, r of $\mathcal{g}(\cdot)$ correspond to $\boldsymbol{O}, \boldsymbol{Z}, \boldsymbol{H}, \boldsymbol{R}$ and indicate the input and output type of $\mathcal{g}(\cdot)$. For example, $\mathcal{g}_{oz}$ denotes a function of $\mathcal{g}(\cdot)$ with the inputs of $\boldsymbol{O}_l, \boldsymbol{O}_{l+1}$ and output of $\boldsymbol{Z}_{l+1}$. σ is the sigmoid function. ι is the leaky Relu function.

As shown in (3), the wave equation [54] in $\mathcal{g}(\cdot)$ describes the relationship of waves (including electromagnetic waves and geoacoustic waves) in time and space,

$$\frac{\partial^2 \boldsymbol{v}}{\partial t^2} = c^2 \nabla^2 \boldsymbol{v} \tag{3}$$

where $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x}, t)$ represents the wave amplitude at position $\boldsymbol{x}$ and time *t*. *c* is the propagation velocity, and ∇^2 is the Laplace operator. The left term describes the variation of the amplitude with time and the right term describes the variation of the amplitude in space.

To integrate the wave equation into GNN block, we replace $\frac{\partial^2 \boldsymbol{v}}{\partial t^2}$ by the temporal variation of node attributes $\left(\boldsymbol{v}_i^{(l+2)} - \boldsymbol{v}_i^{(l+1)}\right) - \left(\boldsymbol{v}_i^{(l+1)} - \boldsymbol{v}_i^{(l)}\right)$, which describes the difference of the node attribute value $\boldsymbol{v}_i$ between *l*, *l*+1 and *l*+2. We use the spatial variation of node attributes $\sum_{j:(i,j)\in\mathcal{E}} w'_{ij}\left(\boldsymbol{v}_i^{(l+1)} - \boldsymbol{v}_j^{(l+1)}\right)$ i.e., the weighted sum of the difference between the node attribute values of $\boldsymbol{v}_i^{l+1}$ and its adjacent nodes, to replace $\nabla^2 \boldsymbol{v}$. $\boldsymbol{w}'_{ij}$ is the weight on the edge from *i* to *j*. Introducing the above two terms into (3), we derive (4).

$$\begin{aligned} &\left(\boldsymbol{v}_i^{(l+2)} - \boldsymbol{v}_i^{(l+1)}\right) - \left(\boldsymbol{v}_i^{(l+1)} - \boldsymbol{v}_i^{(l)}\right) \\ &= \boldsymbol{c}^2 \textstyle\sum_{j:(i,j)\in\mathcal{E}} \boldsymbol{w}'_{ij}\left(\boldsymbol{v}_i^{(l+1)} - \boldsymbol{v}_j^{(l+1)}\right) + \boldsymbol{r}_{ij}^{g} \\ &= \textstyle\sum_{j:(i,j)\in\mathcal{E}} \boldsymbol{w}_{ij}^{g} \boldsymbol{v}_j^{(l+1)} + \boldsymbol{r}_{ij}^{g} \end{aligned} \tag{4}$$

where $\boldsymbol{r}_{ij}^{g}$ is a residual term, which accounts for the deviations caused by the transformation of the theoretical equation. This process is inspired by the loss of the diffusion equation constraints [55].

The right term of (4) is simplified as the weighted sum of the attribute values of node *i* and its adjacent nodes. c^2 and $\boldsymbol{w}'_{ij}$ are combined into $\boldsymbol{w}_{ij}^{g}$, which becomes the weight of the attribute values of node *i* or its adjacent node *j*. Finally, we can derive the update function as (5) where $\boldsymbol{w}_{ij}^{g}, \boldsymbol{r}_{ij}^{g}$ can be learned as optimal values during training. Note $\boldsymbol{v}'_i$ is the vector of the updated node attributes.

$$\boldsymbol{v}'_i = \boldsymbol{v}_i^{(l+1)} + \left(\boldsymbol{v}_i^{(l+1)} - \boldsymbol{v}_i^{(l)}\right) + \textstyle\sum_{j:(i,j)\in\mathcal{E}} \boldsymbol{w}_{ij}^{g} \boldsymbol{v}_j^{(l+1)} + \boldsymbol{r}_{ij}^{g} \tag{5}$$

As shown in Fig.2, the first Physics-informed Recurrent Graph Unit (PRGU) takes $\boldsymbol{O}_1, \boldsymbol{O}_2$ and $\boldsymbol{H}_0, \boldsymbol{H}_1$ as inputs. $\boldsymbol{O}_1, \boldsymbol{O}_2 \in \mathbb{R}^{N\times k}$ represent the observed data on the first and second day of the *t*th week. $\boldsymbol{H}_0$, and $\boldsymbol{H}_1 \in \mathbb{R}^{N\times 32}$ are initialized to two all-zero matrices. The top GNN block $\mathcal{g}(\cdot)$ updates the node attributes based on both spatial neighbors' attributes and previous attributes for observed data $\boldsymbol{O}_1, \boldsymbol{O}_2$. The bottom GNN block $\mathcal{g}(\cdot)$ updates the node attributes for memories $\boldsymbol{H}_0, \boldsymbol{H}_1$.Then the gated units compute the next memory $\boldsymbol{H}_2 \in \mathbb{R}^{N\times 32}$ based on updated observed data and memories. The second PRGU takes $\boldsymbol{O}_2, \boldsymbol{O}_3$ and $\boldsymbol{H}_1, \boldsymbol{H}_2$ as inputs to compute the next memory $\boldsymbol{H}_3$. The operation of PRGU is repeated until all observed data are used. We can use the final memory as the representation for the current observed precursor data.

After obtaining the representations of the observed precursor data, we use a seven-layer fully connected neural network [4] to compute earthquake event representations for the earthquake catalog based on the $\phi(\boldsymbol{q}_t^1, \ldots, \boldsymbol{q}_t^{15})$(i.e. information of the last 15 earthquakes and 8 associated seismic indicators). The information about earthquakes includes magnitude, time of occurrence, epicenter longitude and latitude, and seismic energy. The details of the eight seismic indicators are described in Section IV.A. The earthquake information and indicators are used as the inputs, which have 83 channels. The numbers of hidden channel and output channels are set as 50 and 32, respectively.

The representations of the observed precursor data and the earthquake catalog are fused into the final earthquake event representation $\boldsymbol{e} \in \mathbb{R}^{N\times 64}$ by two fully connected layers. The final event representation is divided into two parts: 1) the magnitude representation, which contains mainly magnitude information, and 2) the epicenter representation, which contains mainly epicenter information. This division helps to find more earthquake representations with similar magnitudes or epicenters, which is useful for earthquake prediction. We add an auxiliary task to predict the magnitude and epicenter based on the magnitude and epicenter representations, respectively. The earthquake pattern-constrained loss of the auxiliary task is of the same form as that of the main task (i.e., the prediction task). The details of the loss function are given in Section III.D.

C. Prior Knowledge Guided Earthquake Prediction Module

The integration of seismic prior knowledge (e.g. earthquake representations and fault distribution) helps to make accurate and physics-informed earthquake predictions. We develop multiple seismic prior knowledge activation layers which serve to activate prior knowledge and event representations in a reciprocal manner, with the goal to generate multi-level activation features for joint magnitude and epicenter prediction. The multi-level activation features consist of activation features extracted from different seismic prior knowledge activation layers, which are concatenated in order to make full use of the different features of the different layers. The seismic prior knowledge used in our work is as follows:

1) Historical earthquake representations. Before prediction, we find similar magnitude and epicenter representations of the current event that is being predicted based on cosine similarity.

2) Spatial distribution of faults [56]. The spatial distribution of faults is processed as gridded data with a spatial resolution of 0.25 °. Each grid takes a value of 0 or 1 to mark the presence or absence of faults in the grid. The height and width of fault data in the study region are 48, 36, respectively.

3) Conditional probability of earthquake recurrence. The conditional probability of earthquake recurrence is computed considering both faults and historical earthquakes. We use the Brownian process time model [57] to calculate the conditional probability of earthquake recurrence for earthquake with

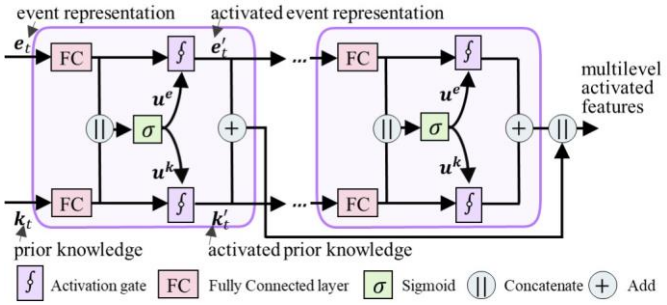


Fig. 3. Seismic prior knowledge activation layers. The event representations activation gate $\boldsymbol{u^e}$ and the prior knowledge activation gate $\boldsymbol{u^k}$ are computed based on the current event representation and prior knowledge and are used to activate the useful information of event representations and prior knowledge respectively. The activated event representations $\boldsymbol{e}'$ and prior knowledge $\boldsymbol{k}'$ will be used as input for the next seismic prior knowledge activation layer and are summed as output. These activation layers output multi-level activated features for the joint forecasting of earthquake magnitude and epicenter.

magnitude above a magnitude threshold for each fault. Based on the difference in activity caused by different faults and different geographical locations, we extend the conditional probability of earthquake recurrence on a fault to the whole study region. p_{f_i} represents the conditional probability of earthquake recurrence of fault f_i. For any location *j* in the study area, the conditional probability of earthquake recurrence p_j is computed as the sum of conditional probabilities on faults weighted by fault activity and activity at *j*,

$$\begin{cases} p_j = \frac{1}{|\mathcal{N}(j)|}\sum_{i\in\mathcal{N}(j)} w^p_{j,f_i} p_{f_i} \\ w^p_{j,f_i} = exp[-\frac{(p_{f_i}-p_{max})^2}{2{\sigma_p}^2} - \frac{{d_{j,f_j}}^2}{2{\sigma_d}^2} - \frac{{r_j}^2}{2{\sigma_r}^2}] \end{cases} \quad (6)$$

where w^p_{j,f_j} is the weight of contribution of f_j to the probability of earthquake recurrence at location *j*, computed by an independent Gaussian function. d_{j,f_j} is the distance between j and f_j. r_j represents the number of historical earthquakes. Using 3.5, 4.5 and 5.5 as the magnitude thresholds, we compute the conditional probability of recurrence for earthquakes with magnitude above 3.5, 4.5 and 5.5, respectively. Thus the size of prior knowledge of the conditional probability of earthquake recurrence is 48×36×3.

We use fully connected layers and Convolutional Neural Network (CNN) [28] to embed historical event representations, grid data of faults and conditional probability of earthquake recurrence in the whole study region into a low-dimension vector $\boldsymbol{k}$ to represent seismic prior knowledge.

After the earthquake event representation $\boldsymbol{e}$ and seismic prior knowledge embedding $\boldsymbol{k}$ are produced, they can be concatenated for earthquake prediction. However, it is not clear which one or which dimension is more important for earthquake prediction. We design prior knowledge activation layers to distill useful information from the earthquake event representation $\boldsymbol{e}$ and seismic prior knowledge embedding $\boldsymbol{k}$ via activation gates.

As shown in Fig. 3, the workflow of the seismic prior knowledge activation layer is as follows: Firstly, the event representations activation gate $\boldsymbol{u^e} \in \mathbb{R}^{N\times 64}$ and the prior knowledge activation gate $\boldsymbol{u^k} \in \mathbb{R}^{N\times 64}$ are computed based on the current event representation $\boldsymbol{e}$ and prior knowledge embedding $\boldsymbol{k}$, corresponding to the first equation in (7). Then, the two activation gates are used to activate the useful information and filter out the useless information of event representations $\boldsymbol{e}$ and prior knowledge embedding $\boldsymbol{k}$ respectively. The activated event representation $\boldsymbol{e}'$ (activated prior knowledge embedding $\boldsymbol{k}'$) is the Hadamard product of $\boldsymbol{e}$ and $\boldsymbol{u^e}$ ($\boldsymbol{k}$ and $\boldsymbol{u^k}$), corresponding to the second (and third equation) in (7). The activated $\boldsymbol{e}'$ and $\boldsymbol{k}'$ will be used as input for the next seismic prior knowledge activation layer. The above steps are repeated for each layer. Finally, multi-level activated features are outputted for earthquake predictions. The above process can be formulated as follows:

$$\begin{cases} \boldsymbol{u^k}, \boldsymbol{u^e} = \sigma(\boldsymbol{W}[\boldsymbol{k}, \boldsymbol{e}] + \boldsymbol{b}) \\ \boldsymbol{k}' = \boldsymbol{u^k}\boldsymbol{k} \\ \boldsymbol{e}' = \boldsymbol{u^e}\boldsymbol{e} \\ \boldsymbol{f}' = \boldsymbol{k}' + \boldsymbol{e}' \end{cases} \quad (7)$$

where $\boldsymbol{W}$ and $\boldsymbol{b}$ are learnable parameters used to compute the gates $\boldsymbol{u^k}, \boldsymbol{u^e}$ based on the current event representation $\boldsymbol{e}$ and the prior knowledge embedding $\boldsymbol{k}$. σ is the sigmoid function.

In our model, we implement and deploy two seismic prior knowledge activation layers. Based on the first layer activation feature $\boldsymbol{f}'$ and the second layer activation feature $\boldsymbol{f}''$, we predict the magnitude and epicenter jointly by two fully connected layers. The number of channels in the last fully connected layer is set to 3 to generate three elements of earthquake simultaneously: magnitude, longitude and latitude of the epicenter.

D. Loss Function

To perform the earthquake magnitude and epicenter forecasting tasks simultaneously, we propose an adaptive multi-task loss function that contains a prediction loss $L_{prediction}$ for the magnitude and epicenter prediction tasks and a pattern constraint loss $L_{constraint}$ for modeling the constraint imposed by magnitude and epicenter patterns. Given that the magnitudes and epicenter distribution are not uniform, a class imbalance problem arises, for which we integrate the probability of earthquake occurrence of a given magnitude and a given location into the prediction loss. We increase the weights of low occurrence examples during training to avoid the clustering of prediction results in areas with high earthquake occurrence.

The loss for the magnitude prediction task is as follows,

$$L_{magnitude} = \frac{1}{n}\sum_{i=1}^{n} f^m(m_i)(m_i - \hat{m}_i)^2 \quad (8)$$

where m_i, $\hat{m}_i$ represent the observed and predicted earthquake magnitudes, respectively. $f^m(\cdot)$ is a mapping function that uses observed magnitude as input to compute the balanced weight for the current sample based on the occurrence probability $p(m_i)$ of earthquakes with magnitude greater than m_i in the study region. $p(m_i)$ is computed by a magnitude probability density function [58], which is based on Gutenberg-Richter's law [35].

$$\begin{cases} f^m(m_i) = \frac{\mu_1}{p(m_i)+\mu_2} \\ p(m_i) = b' \times 10^{-b'(m_i - m_{min})} \end{cases} \quad (9)$$

where $\mu_1, \mu_2 > 0$ are constants we choose. μ_1 is used to ensure that the loss weights are higher than 1 because of the need to amplify the loss of low-probability events. μ_2 is used to avoid the error of dividing by 0. $b' = \frac{b-value}{\log_{10} e}$. m_{min} is the minimum magnitude in the complete earthquake catalog.

Similarly, the loss for the epicenter prediction task can be computed by,

$$L_{epicenter} = \frac{1}{n}\sum_{i=1}^{n} f^{\boldsymbol{y}}(\boldsymbol{y}_i)|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i| \quad (10)$$

where $\boldsymbol{y}_i$, $\hat{\boldsymbol{y}}_i$ represent the observed and forecasted earthquake epicenter vectors consisting of the longitude and latitude of the predicted earthquake, respectively. The norm of $\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i$ is the distance between the observed and predicted earthquake epicenters. $f^{\boldsymbol{y}}(\cdot)$ is a mapping function that computes the balanced weight for the current sample based on the distance to the nearest faults.

$$f^{\boldsymbol{y}}(\boldsymbol{y}_i) = \mu_3 \, exp\left[-\left(\frac{d_{\boldsymbol{y}_i}}{\mu_4}\right)^2\right] + \mu_5 \quad (11)$$

where $\mu_3 < 0, \mu_4, \mu_5 > 0$ are the constants we choose. μ_4 is the bandwidth. μ_5 is used to ensure that the loss weights are higher than 1.

The prediction loss is the sum of the magnitude prediction loss and the epicenter prediction loss,

$$L_{prediction} = L_{magnitude} + exp(-\omega)L_{epicenter} + \lambda\omega \quad (12)$$

where $exp(-\omega)$ is the weight that regulates the two loss contributions of magnitude prediction and epicenter prediction tasks. ω is trained by the adversarial process between $exp(-\omega)$ and $\lambda\omega$. λ is the weight of ω.

The earthquake pattern-constrained loss takes the same form as the prediction losses, except that 'predicted' magnitude and epicenter are computed directly from the magnitude and epicenter representations of earthquake events, rather than multi-level activation features that incorporate seismic prior knowledge. The earthquake pattern-constrained loss is used to drive the useful magnitude and epicenter information being implied in the magnitude and epicenter representations respectively, facilitating the search for similar earthquake representations.

The total loss is the sum of the prediction loss and the earthquake pattern-constrained loss,

$$L_{total} = L_{prediction} + \beta L_{constraint} \quad (13)$$

where β is the weight that regulates the contribution of the two types of losses. During training, the losses are backpropagated to respective modules to train the model.

IV. Data Description and Experimental Settings

A. Data Description

The experiments were conducted on Acoustic & Electromagnetic Testing All in one system (AETA) observation data provided by the Peking University Shenzhen Graduate School [17], and earthquake catalogs provided by China Earthquake Networks Center [59]. The AETA dataset (i.e., observed electromagnetic and geoacoustic data) includes 95 features from 158 stations in the study region from 2017 to 2022. The earthquake catalogs provide information on occurrence time, magnitude, longitude and latitude of the earthquakes. Based on the earthquake catalog data. we calculated eight seismic indicators, including elapsed time, mean magnitude, the rate of square root of seismic energy, slope of the Gutenberg-Richter's law curve (b-value), mean square deviation, magnitude deficit, mean time, coefficient of variation, which are useful indicators for describing earthquakes [22]. The study region is bounded by 22 °–34 ° N and 98 °–107 ° E and has historically experienced a large number of earthquakes. We used four years of data from January 2017 to December 2021 for training, and six months of data from January 2022 to July 2022 for testing. Each training or testing sample consists of observed precursor data$[\boldsymbol{O_1}, \ldots, \boldsymbol{O_l}, \ldots, \boldsymbol{O_7}]_{\boldsymbol{t}} \in \mathbb{R}^{7\times95\times285}$ in the *t* week, features of last 15 earthquakes $\phi(\boldsymbol{q_t^1}, \ldots, \boldsymbol{q_t^{15}}) \in \mathbb{R}^{83}$ before the *t*+1 week, prior knowledge(faults $\in \mathbb{R}^{48\times36}$, conditional probability of earthquake recurrence $\in \mathbb{R}^{48\times36\times3}$& historical event representations $\in \mathbb{R}^{512}$) and a label vector $\boldsymbol{q_{t+1}} \in \mathbb{R}^3$ (the largest earthquake's magnitude, epicenter longitude & latitude in the *t*+1 week). During testing, the proposed STPiDN predicts the largest earthquake in the coming week based on model input of every sample and uses their labels for evaluation.

B. Implementation Details

The prediction model was implemented in python 3.8 using torch1.1.0. All tests were conducted on a desktop machine with an NVIDIA GeForce GTX 2080Ti GPU, a 3.6GHz Intel Core i9-9900K processor, and 32G of memory. The model was trained for 145,000 iterations using Adam with a learning rate of 0.0005. In the first 14,500 iterations, we pre-train the earthquake event representation module (with the earthquake pattern-constrained loss) in order to offer sufficient prior knowledge for subsequent training of the prior knowledge-guided prediction module. Initially, the batch size was set to 5 and then was set to 1 for fine-tuning. We set $\mu_1 = b' + 1$, $\mu_2 = 1$ in (9), $\mu_3 = -1$, $\mu_4 = 50$, $\mu_5 = 2$ in (11), $\lambda = 0.05$ in (12) $\beta = 1$ in (13) respectively because empirically this configuration led to the best prediction results in the tests.

The time costs of the proposed STPiDN (0.956s), physics-informed recurrent graph network (0.187s) and seismic prior knowledge activation layers (0.001s) are not significant for the earthquake prediction task. The losses and parameters converged well during training.

C. Experimental Settings

We use F1 score, Mean Absolute Percentage Error(MAPE) and Mean Distance Error (MDE) to evaluate the prediction performance of the proposed prediction model for earthquake occurrence (magnitude ≥ 3.5), magnitude and epicenter, respectively.

$$F1\_score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (14)$$

where $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$. TP is the number of times the model correctly predicts an upcoming earthquake. TN is the number of times the model does not trigger an alarm and no earthquake occurs. FP is the number of times the model triggers an alarm but no earthquake occurs. FN is the number

of times the model does not trigger an alarm but earthquakes occur. Note that earthquakes with magnitude less than 3.5 are ignored and will not be predicted. Higher F1 scores indicate better performance.

MAPE evaluates the magnitude prediction error when upcoming earthquakes are correctly predicted. MAPE can be calculated as follows,

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{m}_i - m_i|}{m_i} \times 100\% \quad (15)$$

where m_i, $\hat{m}_i$ represent the observed and forecasted earthquake magnitudes, respectively. MAPE is chosen because the absolute error can mitigate the impact of the magnitude imbalance problem on the evaluation.

MDE is the average distance between the observed and predicted epicenters when upcoming earthquakes are correctly predicted,

$$\text{MDE} = \frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i| \quad (16)$$

where $\boldsymbol{y}_i$, $\hat{\boldsymbol{y}}$ represent the observed and predicted earthquake epicenter vectors consisting of longitude and latitude, respectively. The norm of $\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i$ is the distance between the observed and predicted earthquake epicenters.

The performance of the proposed prediction model was compared with five popular baseline models.

- FcNN [4]: The deep Fully connected Neural Network (FcNN) consists of seven fully connected layers and performs well in aftershock earthquake prediction.
- CNN [28]: A network with five convolutional layers and two fully connected layers was used to predict the probability of upcoming earthquakes with magnitude ≥ 4.
- RCNN [31]: A recurrent convolutional neural network was used to predict the occurrence of earthquakes with magnitude above a threshold in grid cells.
- STSGCN [60]: Based on the spatial-temporal synchronous modeling mechanism, the spatial-temporal synchronous graph convolutional networks capture localized spatial-temporal correlations and heterogeneity for spatial-temporal network data prediction.
- ConvLSTM [61]: A multitask ConvLSTM Encoder-Decoder, which was originally used for crowd density and crowd volume prediction tasks, was used for the prediction of earthquake occurrence (magnitude ≥ 3.5), magnitude and epicenter in this study.

These compared baselines involve a variety of commonly-used deep learning techniques, such as NN, CNN, GNN, and ConvLSTM. Three of compared baselines (i.e., FcNN, CNN and RCNN) were originally designed for earthquake prediction and performed well in earthquake prediction[4], [28], [31]. These baselines can be readily used for spatio-temporal event prediction and thus were selected for performance comparison in the study. In order to keep the inputs consistent across models, earthquake precursor data and prior knowledge (fault distribution & conditional probability of earthquake recurrence) were also fed into the baselines after being processed into a suitable form.

In addition to the baseline models described above, we also performed an ablation study by evaluating the prediction performance of four variants of the proposed model.

- STPiDN -LOSS: This model considers only the MSE loss.
- STPiDN -PRGU: This model replaces the PRGU with a vanilla recurrent graph network.
- STPiDN -PKAL: This model replaces the seismic prior knowledge activation layers with a fully connected layer.
- STPiDN -JointF: This model does not perform multi-task prediction and performs only magnitude prediction.

TABLE I

COMPARISON OF EARTHQUAKE OCCURRENCE, MAGNITUDE AND EPICENTER FORECASTING PERFORMANCE. THE TEST WAS REPEATED FIVE TIMES. THE MEAN AND STANDARD DEVIATIONS OF THE EVALUATION METRICS ARE REPORTED. -JOINTF DOES NOT HAVE RESULTS FOR MDE BECAUSE IT DOES NOT PERFORM EPICENTER PREDICTIONS.

Method	F1 score	Precision	Recall	MAPE(%)	MDE(km)
FcNN	0.719±0.030	0.592±0.063	0.938±0.090	16.3±1.4	387± 45
CNN	0.684±0.000	0.520±0.000	1.000±0.000	13.8±2.8	379± 77
RCNN	0.706±0.035	0.554±0.057	0.985±0.031	13.4±1.7	386± 13
STSGCN	0.707±0.023	0.548±0.027	**1.000**±0.000	18.5±4.5	454±148
ConvLSTM	0.753±0.023	0.631±0.034	0.938±0.058	16.1±1.4	385± 42
-LOSS	0.746±0.023	0.607±0.030	0.969±0.038	15.1±1.4	413± 36
-PRGU	0.763±0.022	0.624±0.035	0.985±0.031	15.7±4.9	442±111
-PKAL	0.758±0.023	0.635±0.060	0.954±0.062	14.7±2.2	503± 60
-JointF	0.716±0.020	0.570±0.028	0.969±0.038	**11.1**±1.3	/
STPiDN	**0.778**±0.020	**0.666**±0.036	0.938±0.031	13.5±2.9	**328**± 47

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. Comparison with Baselines

We compared the performance of STPiDN with the baselines and its four variants in forecasting the largest earthquake in the study area in the coming week. Table I shows the evaluation results for earthquake occurrence, magnitude and epicenter prediction performance. Compared with baselines, STPiDN provides a better F1 score, MAPE and MDE, and small standard deviations of these metrics. This indicates that the proposed STPiDN shows a good and stable performance in earthquake prediction. The F1 score of ConvLSTM is higher than other baselines, which is probably caused by the multitask learning used in ConvLSTM. STSGCN does not perform well in MAPE and MDE because earthquake forecasting is more challenging compared to traffic prediction due to its stochastic and complex physical processes. The STPiDN outperforms FcNN by a large margin because the STPiDN integrates observed precursor data and seismic prior knowledge in a physics-informed manner.

B. Ablation Study

Table I also shows the results of the ablation study. We evaluated the contribution of four components in STPiDN: the physics-informed recurrent graph units, earthquake prior knowledge activation layers, the adaptive loss function and the joint magnitude and epicenter forecasting scheme. The evaluation of these components corresponds –PGRU, -PKAL, -LOSS and –JointF. The differences between –JointF and –LOSS are that: 1) –LOSS still forecasts magnitude and epicenter jointly, but uses the MSE loss in all tasks; 2) –JointF uses only the magnitude prediction loss and earthquake pattern constrained loss for magnitude prediction.

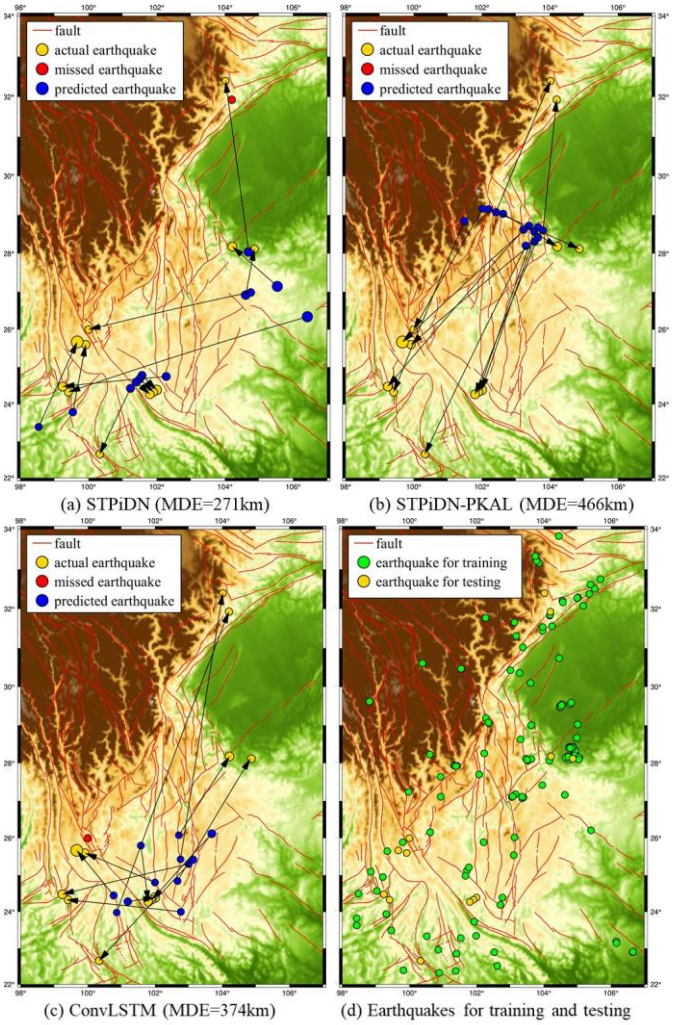


Fig. 4. Comparison of weekly predicted earthquakes and actual earthquakes from January 7, 2021 to July 14, 2021. The topographic map was produced using the data provided by the Shuttle Radar Topography Mission [62]. Yellow dots represent actual earthquakes that are correctly predicted. Red dots represent missed earthquakes. Blue dots represent predicted earthquakes. Black arrows point from the predicted earthquakes to the corresponding actual earthquakes. The dots are slightly offset to avoid visual occlusions. The MDE is the average distance between the observed and predicted epicenters. The earthquakes for training (represented by yellow dots in (d)) were collected from January 2017 to December 2021, while the earthquakes for testing (represented by green dots in (d)) were collected from January 2022 to July 2022.

–JointF provides the lowest F1 score among the four variants, indicating that joint magnitude and epicenter forecasting performs better than single-task forecasting because magnitude and epicenter forecasting tasks are related and can complement each other for promoting forecasting performance. Compared to other variants, the F1 score of –JointF decreases most significantly, indicating that joint forecasting plays the most important role in STPiDN. The F1 score of –PRGU is slightly lower than that of STPiDN, suggesting the integration of wave equation in the update function of GNN blocks plays an important but minor role in earthquake occurrence forecasting. The poor performance of MDE of –PKAL may be caused by the insufficient integration of prior knowledge, such as fault distribution that is useful for epicenter prediction. –JointF provides the lowest MAPE.

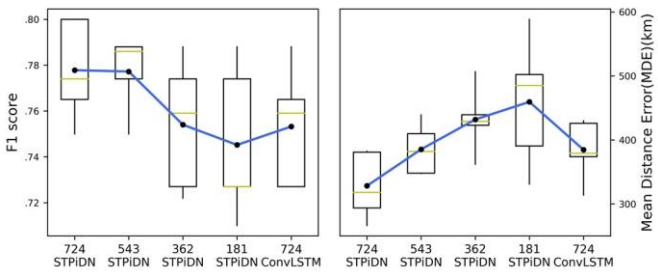


Fig. 5. Comparison of earthquake prediction performance of STPiDN on training samples of different sizes. The black dots indicate the average values of F1 score and MDE for 5 runs.

The ablation study shows that: 1) In terms of earthquake occurrence prediction performance, the most useful component is the multi-task learning scheme while the physics-informed recurrent graph units are the least important; 2) the integration of prior knowledge plays an important role in earthquake epicenter prediction; and 3) the components of STPiDN complement each other to promote earthquake magnitude and epicenter prediction performance.

C. Visual Analytics of Forecasted Earthquakes

In addition to the quantitative performance metrics, we also made a visual comparison of actual and predicted epicenters in the study region. Fig. 4 shows the weekly predicted epicenters of STPiDN and its variant –PKAL with the corresponding actual epicenters from January 7, 2021 to July 14, 2021 (total 25 weeks). Earthquakes (magnitude 3.5 or greater) occurred in 13 of the 25 weeks with no earthquakes in the remaining 12 weeks.

It can be observed from Fig. 4(a) that: 1) most of the earthquake events can be correctly predicted by STPiDN; and 2) most of the predicted epicenters are within 300km of the actual epicenters, especially in the southwest where some of the offsets are within 100km. These observations show that our method is able to capture some occurrence patterns of earthquakes. There is one missed earthquake and three earthquakes with epicenters offset over 400km. To find out the reason why STPiDN made mistakes on these events, we analyzed similar event representations with these events. We found that similar event representations of the missed earthquake are almost from events without earthquake occurrence. This means that the representation of the missed earthquake is not close to earthquake events in the latent space. The reason for large epicenter offsets is that it is difficult to find similar epicenter representations due to the spatial sparsity of earthquakes. The prediction performance can be further promoted by improving the earthquake event representations. As can be seen in Fig. 4(b), without PKAL, the predicted epicenters tend to be concentrated in the central region, indicating that PKAL can improve the localization of epicenter

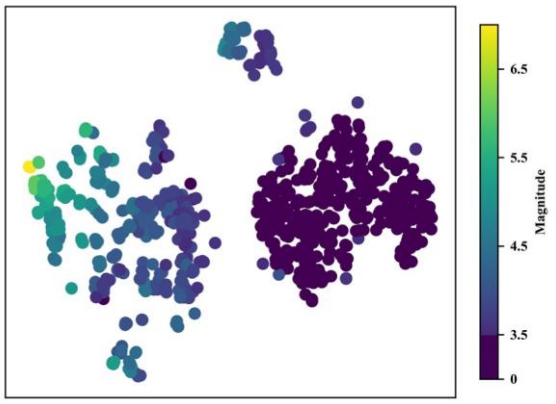


Fig. 6. t-SNE visualization of the earthquake event representations learned by STPiDN. Events are represented as dots, colored according to their magnitude.

of future earthquakes by the effective integration of prior knowledge.

D. Evaluation of The Effect of Sample Sizes

We compared the prediction performance of the proposed STPiDN on different sizes of training samples with the best baseline to evaluate the generalization capability of STPiDN. Fig. 5 shows the performance of STPiDN on training samples of different sizes (724, 543, 362 and 181 training samples), which include 206, 161, 120 and 50 earthquakes, respectively. When the sample size is reduced from 724 to 543, the average F1 score of STPiDN barely decreases. When the sample size is reduced to 362, STPiDN still provides a higher average F1 score than the best baseline ConvLSTM trained on 724 samples, suggesting that STPiDN can perform well in earthquake occurrence prediction even with small training samples due to its relatively good generalization capability. The MDE values decrease rapidly as the training sample size decreases, indicating that training on fewer training samples has a greater adverse impact on earthquake epicenter prediction. However, the average MDE of STPiDN trained on 543 training samples is only 0.8km lower than the best baseline ConvLSTM trained on 724 training samples. The above results suggest that STPiDN can perform well in earthquake prediction with small training samples, which may be due to the integration of prior knowledge that reduces the need for training samples. STPiDN can leverage similar earthquake magnitude and epicenter representations to better capture the spatio-temporal occurrence patterns of earthquakes, thereby reducing the need for a large number of training samples.

E. Visual Analytics of Earthquake Event Representations

We visualized earthquake event representations learned by STPiDN. Fig. 6 provides a t-SNE visualization of the earthquake event representations. Fig. 6 shows that: 1) There is a clear distinction between the latent representations of earthquake events occurrences (earthquake with magnitude above 3.5) and representations without earthquake occurrences; 2) Earthquakes with large magnitudes are separated from those with small magnitudes; and 3) Representations of larger earthquakes are more distant from the representations without earthquake occurrences than those of smaller earthquakes. These findings indicate the proposed STPiDN can embed useful information into earthquake magnitude representation, which helps the improvement of earthquake occurrence and magnitude prediction performance.

VI. Conclusion

We have introduced a data-driven approach to jointly predict earthquake magnitude and epicenter in a physics-informed manner. In the proposed STPiDN, we develop a physics-informed recurrent graph network based on wave equation to represent historical seismic events. We develop prior knowledge activation layers to fuse the seismic prior knowledge with earthquake event representation. We use the adaptive multitask loss to achieve joint magnitude and epicenter forecasting, with the benefit of alleviating the magnitude & epicenter imbalance problem. STPiDN outperforms several existing methods on a real-world earthquake dataset.

In our study, we found that the alleviation of magnitude/ epicenter imbalance is critical for accurate earthquake prediction. Uneven magnitude/epicenter distribution problem should be considered in future earthquake prediction in terms of data processing or loss function design. Another important research topic is how models can predict large earthquakes when they have not been observed in the training dataset. We think leveraging the similarity of historical earthquakes can help address this problem.

There is still much work to be done for short-term earthquake prediction. From the perspective of data, high-quality precursor datasets should be created. For example, when arranging the location of stations, we need to consider both the locations of the stations and the seismicity characteristics of different locations to obtain sufficient and spatially-balanced observation data. From the perspective of modelling, advanced prior knowledge informed data-driven models should be developed. More seismic prior knowledge can be adequately considered, such as Bath's law [62] and Omori's law [63]. More efforts can be put into designing specific deep learning model structures that incorporate seismic prior knowledge, such as equivariance-preserving deep spatial transformer [49] and physics-informed recurrent graph network. We believe the integration of seismic prior knowledge is beneficial for the prediction of earthquakes and the proposed PRGN is possible to be applied in other regions if the required seismic data is available for these regions. This topic can be explored in future research.

Moreover, focal depth is worth exploring in data driven earthquake forecasting because it indicates the tectonic background of earthquakes and seismicity patterns. Focal depths and epicenters encode spatial information of earthquakes. The spatial distribution pattern analyzed from historical focal depths and epicenters can facilitate earthquake prediction and prevention. Moreover, intensity represents the extent of damage caused by earthquakes and can be calculated based on focal depth. In the current model, we did not explicitly account for the focal depth information. The multitask learning scheme used in the current model (magnitude prediction task, epicenter prediction task and event representation constraint task) is overly complicated. The focal depth prediction task was not

considered in the current model due to the model convergence problem (e.g. overfitting) caused by the complex model structure. In the future work, we will develop a parsimonious model that incorporates the focal depth prediction subtask.

ACKNOWLEDGMENT

The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

APPENDIX

A. Details of pre-processing and entries of the training dataset

The observed electromagnetic and geoacoustic data provided by AETA includes 95 features per 10 minutes from 158 stations from 2017 to 2022. The electromagnetic and geoacoustic data are stored in EXCEL format by week and station. We read these data in python and concatenated all the data together to form a tensor with a size of Time×Station_number×Feature_number. We excluded stations that were not operational continuously from 2017 to 2022, or with severe data missing problems. The final number of stations used in the study was 95. A week of data with a temporal resolution of 10 minutes forms a tensor with a size of 1008×95×95(Time=7×24×6=1008). To simplify the input, we only kept the mean, maximum and minimum values for each feature for each day. Thus, the time dimension was simplified to 7 days, while the feature dimension became 95×3=285. Finally, the size of the input, i.e. weekly observed precursor tensor $[\boldsymbol{O_1}, \dots, \boldsymbol{O_l}, \dots, \boldsymbol{O_7}]_t$ is 7×95×285. The size of $\boldsymbol{O_l}$, i.e., the data on the *l*th day of the *t*th week, is 95×285. The proposed Physics-informed Recurrent Graph Network (PRGN) literately processes $\boldsymbol{O_l}$ by Physics-informed Recurrent Graph Unit(PRGU) to produce seismic embeddings from observed precursor data $[\boldsymbol{O_1}, \dots, \boldsymbol{O_l}, \dots, \boldsymbol{O_7}]_t$.

The entries of the training dataset include (1) observed precursor tensor $[\boldsymbol{O_1}, \dots, \boldsymbol{O_l}, \dots, \boldsymbol{O_7}]_t$ in the *t* week, (2) features of last 15 earthquakes $\phi(\boldsymbol{q_t^1}, \dots, \boldsymbol{q_t^{15}})$ before the *t*+1 week, (3) prior knowledge and (4) label vector $\boldsymbol{q_{t+1}}$ (the largest earthquake's magnitude, epicenter longitude & latitude in the *t*+1 week).

(1) Observed precursor tensor $[\boldsymbol{O_1}, \dots, \boldsymbol{O_l}, \dots, \boldsymbol{O_7}]_t \in \mathbb{R}^{7\times95\times285}$. The tensor size of the observed precursor is 7×95×285. 7 represents the number of days of inputs. 95 is the number of stations of inputs. 285 is the number of features for each station, corresponding to the mean, maximum and minimum values for one day of 95 features provided by AETA. The observed precursor tensor is embedded into event representation vectors by PRGN.

(2) Features of the last 15 earthquakes $\phi(\boldsymbol{q_t^1}, \dots, \boldsymbol{q_t^{15}}) \in \mathbb{R}^{83}$ consist of 5 basic information of the last 15 earthquakes and 8 associated seismic indicators. The 5 basic information about earthquakes includes magnitude, time of occurrence, epicenter longitude and latitude, and seismic energy. The 8 seismic indicators are elapsed time, mean magnitude, the rate of square root of seismic energy, slope of the Gutenberg-Richter's law curve (b-value), mean square deviation, magnitude deficit, mean time, and coefficient of variation. The size of $\phi(\boldsymbol{q_t^1}, \dots, \boldsymbol{q_t^{15}})$ (83) is the sum of the number of basic information of the last 15 earthquakes and the number of seismic indicators ($5 \times 15+8$). The features of the last 15 earthquakes are fused into event representation vectors by fully connected layers.

(3) Prior knowledge includes a fault map $\in \mathbb{R}^{48\times36}$, a conditional probability of earthquake recurrence $\in \mathbb{R}^{48\times36\times3}$ and historical event representations $\in \mathbb{R}^{512}$. The first two priors are gridded data with a size of 48×36. Each grid in the fault map takes a value of 0 or 1 to mark the presence or absence of faults in the grid. The conditional probability of earthquake recurrence represents the conditional probability of recurrence for earthquakes with magnitude above 3.5, 4.5, and 5.5 in each grid. The historical event representations are the concatenation of eight event representations that are similar to the current event representation. The size of each event representation is 64. The size of historical event representations is 512 (i.e. 64×8). These data are fused into prior knowledge embedding by convolutional neural network and fully connected layers.

(4) Label vector $\boldsymbol{q_{t+1}} \in \mathbb{R}^3$ includes the magnitude, longitude & latitude of epicenter of the largest earthquake in the *t*+1 week. The label vector $\boldsymbol{q_{t+1}}$ is used as ground truth during training and evaluation.

REFERENCES

[1] W. Zhou, Y. Liang, Z. Ming, and H. Dong, "Earthquake Prediction Model Based on Danger Theory in Artificial Immunity," *NNW*, vol. 30, no. 4, pp. 231–247, 2020, doi: 10.14311/NNW.2020.30.016.

[2] R. Li, X. Lu, S. Li, H. Yang, J. Qiu, and L. Zhang, "DLEP: A Deep Learning Model for Earthquake Prediction," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9207621.

[3] R. Yuan, "An improved K-means clustering algorithm for global earthquake catalogs and earthquake magnitude prediction," *Journal of Seismology*, vol. 25, no. 3, pp. 1005–1020, 2021.

[4] P. M. R. DeVries, F. Viégas, M. Wattenberg, and B. J. Meade, "Deep learning of aftershock patterns following large earthquakes," *Nature*, vol. 560, no. 7720, Art. no. 7720, Aug. 2018, doi: 10.1038/s41586-018-0438-y.

[5] G. C. Beroza, M. Segou, and S. Mostafa Mousavi, "Machine learning and earthquake forecasting—next steps," *Nat Commun*, vol. 12, no. 1, Art. no. 1, Aug. 2021, doi: 10.1038/s41467-021-24952-6.

[6] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification," *Soil Dynamics and Earthquake Engineering*, vol. 111, pp. 1–7, Aug. 2018, doi: 10.1016/j.soildyn.2018.04.020.

[7] T. Chelidze, G. Melikadze, T. Kiria, T. Jimsheladze, and G. Kobzev, "Statistical and Non-linear Dynamics Methods of Earthquake Forecast: Application in the Caucasus," *Front. Earth Sci.*, vol. 0, 2020, doi: 10.3389/feart.2020.00194.

[8] G. Asencio–Cortés, A. Morales–Esteban, X. Shang, and F. Martínez–Álvarez, "Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure," *Computers & Geosciences*, vol.

115, pp. 198–210, Jun. 2018, doi: 10.1016/j.cageo.2017.10.011.

[9] R. Shcherbakov, J. Zhuang, G. Zöller, and Y. Ogata, "Forecasting the magnitude of the largest expected earthquake," *Nat Commun*, vol. 10, no. 1, Art. no. 1, Sep. 2019, doi: 10.1038/s41467-019-11958-4.

[10] H. O. Cekim, S. Tekin, and G. Özel, "Prediction of the earthquake magnitude by time series methods along the East Anatolian Fault, Turkey," *Earth Sci Inform*, vol. 14, no. 3, pp. 1339–1348, Sep. 2021, doi: 10.1007/s12145-021-00636-z.

[11] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical association*, vol. 83, no. 401, pp. 9–27, 1988.

[12] T. Chetia, S. Baruah, C. Dey, S. Sharma, and S. Baruah, "Probabilistic analysis of seismic data for earthquake forecast in North East India and its vicinity.," *Current Science (00113891)*, vol. 117, no. 7, 2019.

[13] Y. Xie, M. Ebad Sichani, J. E. Padgett, and R. DesRoches, "The promise of implementing machine learning in earthquake engineering: A state-of-the-art review," *Earthquake Spectra*, vol. 36, no. 4, pp. 1769–1801, 2020.

[14] A. Mignan and M. Broccardo, "Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations," *Seismological Research Letters*, vol. 91, no. 4, pp. 2330–2342, 2020.

[15] F. Khosravikia, P. Clayton, and Z. Nagy, "Artificial neural network-based framework for developing ground-motion models for natural and induced earthquakes in Oklahoma, Kansas, and Texas," *Seismological Research Letters*, vol. 90, no. 2A, pp. 604–613, 2019.

[16] A. Khalatbarisoltani, M. Soleymani, and M. Khodadadi, "Online control of an active seismic system via reinforcement learning," *Structural Control and Health Monitoring*, vol. 26, no. 3, p. e2298, 2019, doi: 10.1002/stc.2298.

[17] Wang, Xinan *et al.*, "Research and Implementation of Multi-component Seismic Monitoring System AETA," *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 54, no. 3, pp. 487–494, 2018, doi: 10.13209/j.0479-8023.2017.171.

[18] Y. Liu *et al.*, "An Earthquake Forecast Model Based on Multi-Station PCA Algorithm," *Applied Sciences*, vol. 12, no. 7, p. 3311, 2022.

[19] Z. Bao, J. Zhao, P. Huang, S. Yong, and X. Wang, "A deep learning-based electromagnetic signal for earthquake magnitude prediction," *Sensors*, vol. 21, no. 13, p. 4434, 2021.

[20] R. Mallouhy, C. A. Jaoude, C. Guyeux, and A. Makhoul, "Major earthquake event prediction using various machine learning algorithms," in *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Dec. 2019, pp. 1–7. doi: 10.1109/ICT-DM47966.2019.9032983.

[21] S. Karimzadeh, M. Matsuoka, J. Kuang, and L. Ge, "Spatial Prediction of Aftershocks Triggered by a Major Earthquake: A Binary Machine Learning Perspective," *ISPRS International Journal of Geo-Information*, vol. 8, no. 10, Art. no. 10, Oct. 2019, doi: 10.3390/ijgi8100462.

[22] A. Panakkat and H. Adeli, "Neural Network Models for Earthquake Magnitude Prediction Using Multiple Seismicity Indicators," *International Journal of Neural Systems*, vol. 17, no. 1, pp. 13–33, 2007, doi: 10.1142/S0129065707000890.

[23] A. S. Alarifi, N. S. Alarifi, and S. Al-Humidan, "Earthquakes magnitude predication using artificial neural network in northern Red Sea area," *Journal of King Saud University-Science*, vol. 24, no. 4, pp. 301–313, 2012.

[24] S. Asaly, L.-A. Gottlieb, N. Inbar, and Y. Reuveni, "Using Support Vector Machine (SVM) with GPS Ionospheric TEC Estimations to Potentially Predict Earthquake Events," *Remote Sensing*, vol. 14, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/rs14122822.

[25] J. Zhu, S. Li, and J. Song, "Magnitude estimation for earthquake early warning with multiple parameter inputs and a support vector machine," *Seismological Research Letters*, vol. 93, no. 1, pp. 126–136, 2022.

[26] A. Galkina and N. Grafeeva, "Machine learning methods for earthquake prediction: A survey," in *Proceedings of the Fourth Conference on Software Engineering and Information Management (SEIM-2019), Saint Petersburg, Russia*, 2019, p. 25.

[27] M. Yousefzadeh, S. A. Hosseini, and M. Farnaghi, "Spatiotemporally explicit earthquake prediction using deep neural network," *Soil Dynamics and Earthquake Engineering*, vol. 144, p. 106663, May 2021, doi: 10.1016/j.soildyn.2021.106663.

[28] R. Jena, B. Pradhan, S. P. Naik, and A. M. Alamri, "Earthquake risk assessment in NE India using deep learning and geospatial analysis," *Geoscience Frontiers*, vol. 12, no. 3, p. 101110, May 2021, doi: 10.1016/j.gsf.2020.11.007.

[29] J. Huang, X. Wang, Y. Zhao, C. Xin, and H. Xiang, "Large earthquake magnitude prediction in Taiwan based on deep learning neural network," *Neural Network World*, vol. 28, no. 2, pp. 149–160, 2018.

[30] Y. Pu, J. Chen, and D. B. Apel, "Deep and confident prediction for a laboratory earthquake," *Neural Comput & Applic*, vol. 33, no. 18, pp. 11691–11701, Sep. 2021, doi: 10.1007/s00521-021-05872-4.

[31] R. Kail, E. Burnaev, and A. Zaytsev, "Recurrent Convolutional Neural Networks Help to Predict Location of Earthquakes," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3107998.

[32] S. M. Mousavi and G. C. Beroza, "Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–14, 2020.

[33] N. C. Ristea and A. Radoi, "Complex Neural Networks for Estimating Epicentral Distance, Depth, and Magnitude of Seismic Waves," *IEEE geoscience and remote sensing letters*, no. 19-, 2022.

[34] A. Kato and Y. Ben-Zion, "The generation of large earthquakes," *Nat Rev Earth Environ*, vol. 2, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s43017-020-00108-w.

[35] B. Gutenberg and C. F. Richter, "Frequency of Earthquakes in California," *Bulletin of the Seismological Society of America*, vol. 34, no. 4, pp. 185–188, 1994.

[36] S. Pulinets, "Ionospheric Precursors of Earthquakes; Recent Advances in Theory and Practical Applications," *Terrestrial Atmospheric & Oceanic Sciences*, vol. 15, no. 3, pp. 413–435, 2004.

[37] V. Gitis, A. Derendyaev, and K. Petrov, "Analyzing the Performance of GPS Data for Earthquake Prediction," *Remote Sensing*, vol. 13, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/rs13091842.

[38] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2015.

[39] H. Yao *et al.*, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2018.

[40] A. Ali, Y. Zhu, and M. Zakarya, "Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction," *Neural Networks*, vol. 145, pp. 233–247, Jan. 2022, doi: 10.1016/j.neunet.2021.10.021.

[41] O. Medjaouri and K. Desai, "HR-STAN: High-Resolution Spatio-Temporal Attention Network for 3D Human Motion Prediction," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2540–2549.

[42] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very Deep Convolutional Networks for Text Classification." arXiv, Jan. 27, 2017. doi: 10.48550/arXiv.1606.01781.

[43] L. von Rueden *et al.*, "Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021, doi: 10.1109/TKDE.2021.3079836.

[44] S. Seo, C. Meng, and Y. Liu, "Physics-aware Difference Graph Networks for Sparsely-Observed Dynamics," presented at the International Conference on Learning Representations, Sep. 2019.

[45] Y. Yin *et al.*, "Augmenting physical models with deep networks for complex dynamics forecasting," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2021, no. 12, p. 124012, 2021.

[46] Y. Lu, M. Rajora, P. Zou, and S. Y. Liang, "Physics-Embedded Machine Learning: Case Study with Electrochemical Micro-Machining," *Machines*, vol. 5, no. 1, Art. no. 1, Mar. 2017, doi: 10.3390/machines5010004.

[47] E. de Bézenac, A. Pajot, and P. Gallinari, "Deep learning for physical processes: incorporating prior scientific knowledge," *J. Stat. Mech.*, vol. 2019, no. 12, p. 124009, 2019, doi: 10.1088/1742-5468/ab3195.

[48] J. Ling, A. Kurzawski, and J. Templeton, "Reynolds averaged turbulence modelling using deep neural networks with embedded invariance," *Journal of Fluid Mechanics*, vol. 807, pp. 155–166, 2016.

[49] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, and K. Kashinath, "Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving deep spatial transformers." arXiv, Mar. 16, 2021. doi: 10.48550/arXiv.2103.09360.

[50] J. S. Read *et al.*, "Process-guided deep learning predictions of lake water temperature," *Water Resources Research*, vol. 55, no. 11, pp. 9173–9190, 2019.

[51] S. Esmaeilzadeh *et al.*, "Meshfreeflownet: A physics-constrained deep continuous space-time super-resolution framework," in *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, IEEE, 2020, pp. 1–15.

[52] C. Brunsdon, S. Fotheringham, and M. Charlton, "Geographically weighted regression," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, no. 3, pp. 431–443, 1998.

[53] Y. Li, R. Zemel, M. Brockschmidt, and D. Tarlow, "Gated Graph Sequence Neural Networks," presented at the Proceedings of ICLR'16, Apr. 2016.

[54] M. Renardy and R. C. Rogers, *An introduction to partial differential equations*, vol. 13. Springer Science & Business Media, 2006.

[55] S. Seo and Y. Liu, "Differentiable Physics-informed Graph Networks." arXiv, Feb. 10, 2019. doi: 10.48550/arXiv.1902.02950.

[56] Q. D. Deng, P. Z. Zhang, Y. K. Ran, X. P. Yang, M. Wei, and L. C. Chen, "Active Tectonics and Earthquake Activities in China," *Earth Science Frontiers*, 2003.

[57] W. L. Ellsworth, M. V. Matthews, R. M. Nadeau, S. P. Nishenko, P. A. Reasenberg, and R. W. Simpson, "A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities," *US Geological Survey Open-File Report*, vol. 99, no. 522, p. 22, 1999.

[58] A. De Santis, G. Cianchini, P. Favali, L. Beranzoli, and E. Boschi, "The Gutenberg–Richter Law and Entropy of Earthquakes: Two Case Studies in Central Italy," *Bulletin of the Seismological Society of America*, vol. 101, no. 3, pp. 1386–1395, Jun. 2011, doi: 10.1785/0120090390.

[59] "China Earthquake Networks Center. (2006, Jan.). China Earthquake Networks Center. [Online]. Available: https://news.ceic.ac.cn (In Chinese)."

[60] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, Art. no. 01, Apr. 2020, doi: 10.1609/aaai.v34i01.5438.

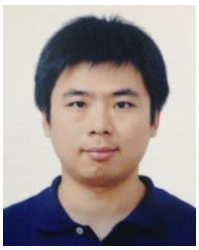
[61] R. Jiang, X. Song, D. Huang, X. Song, and R. Shibasaki, "DeepUrbanEvent: A System for Predicting Citywide Crowd Dynamics at Big Events," in *the 25th ACM SIGKDD International Conference*, 2019.

[62] B. Tozer, D. T. Sandwell, W. H. Smith, C. Olson, J. Beale, and P. Wessel, "Global bathymetry and topography at 15 arc sec: SRTM15+," *Earth and Space Science*, vol. 6, no. 10, pp. 1847–1864, 2019.

[63] F. Omori, "On the after-shocks of earthquakes," PhD Thesis, The University of Tokyo, 1895.

Jie Liu received the B.Eng. degree from the China University of Petroleum, Qingdao, China, in 2018. She is pursuing Doctor's Degree with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuwan University, Wuhan, China.

Her research interests include spatio-temporal prediction and deep learning.



Tong Zhang received the M.Eng. degree in cartography and geographic information system (GIS) from Wuhan University, Wuhan, China, in 2003, and the Ph.D. degree in geography from San Diego State University, San Diego, CA, and the University of California at Santa Barbara, Santa Barbara, CA, in 2007.

He is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research topics include spatio-temporal machine learning, high-resolution remote sensing image analysis, and urban computing.

Chulin Gao received the B.S. degree and the M.S. degree in geophysics from Wuhan University, Wuhan, China, in 2019 and 2022 respectively. She is pursuing Doctor's Degree with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

Her research interests include spatio-temporal prediction, coseismic water level response mechanism and earthquake precursor research.

Peixiao Wang is a PhD candidate for State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He received the M.S. degree from The Academy of Digital China, Fuzhou University in 2020. In the past years, he has published over 10 refereed journal articles and conference papers as the first or corresponding author.

His researches focus on spatio-temporal data mining, social computing, and public health.